

LONDON  
SCHOOL of  
HYGIENE  
& TROPICAL  
MEDICINE



Pickering, H; Teng, A; Faal, N; Joof, H; Makalo, P; Cassama, E; Nabicassa, M; Last, AR; Burr, SE; Rowland-Jones, SL; Thomson, NR; Roberts, CH; Mabey, DCW; Bailey, RL; Hayward, RD; de la Maza, LM; Holland, MJ (2017) Genome-wide profiling of humoral immunity and pathogen genes under selection identifies immune evasion tactics of *Chlamydia trachomatis* during ocular infection. *Sci Rep*, 7 (1). p. 9634. ISSN 2045-2322 DOI: <https://doi.org/10.1038/s41598-017-09193-2>

Downloaded from: <http://researchonline.lshtm.ac.uk/4327482/>

DOI: [10.1038/s41598-017-09193-2](https://doi.org/10.1038/s41598-017-09193-2)

#### Usage Guidelines

Please refer to usage guidelines at <http://researchonline.lshtm.ac.uk/policies.html> or alternatively contact [researchonline@lshtm.ac.uk](mailto:researchonline@lshtm.ac.uk).

Available under license: <http://creativecommons.org/licenses/by/2.5/>

# SCIENTIFIC REPORTS

OPEN

## Genome-wide profiling of humoral immunity and pathogen genes under selection identifies immune evasion tactics of *Chlamydia trachomatis* during ocular infection

Harry Pickering<sup>1</sup>, Andy Teng<sup>2</sup>, Nkoyo Faal<sup>3</sup>, Hassan Joof<sup>3</sup>, Pateh Makalo<sup>3</sup>, Eunice Cassama<sup>4</sup>, Meno Nabicassa<sup>4</sup>, Anna R. Last<sup>1</sup>, Sarah E. Burr<sup>1,3</sup>, Sarah L. Rowland-Jones<sup>3</sup>, Nicholas R. Thomson<sup>5,6</sup>, Chrissy h. Roberts<sup>1</sup>, David C. W. Mabey<sup>1</sup>, Robin L. Bailey<sup>1</sup>, Richard D. Hayward<sup>7</sup>, Luis M. de la Maza<sup>8</sup> & Martin J. Holland<sup>1,3</sup>

The frequency and duration of *Chlamydia trachomatis* (Ct) ocular infections decrease with age, suggesting development of partial immunity. However, there is a lack of clear correlates of immunity to Ct infection in humans. We screened sera from a cohort of Gambian children followed for six-months against a Ct-proteome microarray. At genome sequence level, we detected signatures of selection from a population of ocular Ct isolates from Guinea-Bissau. Together these approaches allowed us to highlight the focus of humoral responses and hypothesise new modes of pathogen immune evasion. Children who were susceptible to frequent and/or prolonged Ct infection had a less focussed antibody response, including preferential recognition of forty-two antigens. There was evidence of positive and purifying selection across the genome, but little balancing selection. In contrast, most antigens that were associated with susceptibility were under neutral selection. These data suggest an evasion strategy in which Ct presents a large panel of irrelevant antigens to the immune system to block or misdirect protective responses. Development of a focused immune response, possibly induced through vaccination, may be an effective strategy to promote protection to Ct infection.

Ocular infection with *Chlamydia trachomatis* (Ct) causes trachoma, the leading infectious cause of blindness<sup>1</sup>. Both ocular Ct infection and active disease prevalence decline from their peaks in pre-school children (one to four years old) to older children (five to fourteen years old) and from this group to adults (fifteen years or older)<sup>2,3</sup>. This suggests that partial Ct immunity develops with increasing age in endemic communities, notwithstanding reduced exposure to Ct with increasing age<sup>4</sup>.

Conjunctival Ct infection induces a strong pro-inflammatory response marked by production of cytokines<sup>5</sup>, recruitment of neutrophils, macrophages and NK-cells<sup>6,7</sup>. Induction and proliferation of CD4<sup>+</sup> T-cells and production of interferon-gamma (IFN $\gamma$ ) have been implicated in successful resolution of infection in animal models and human infections<sup>8–11</sup>. Ocular Ct infection additionally induces local and systemic antibodies<sup>12</sup>. Neutralising antibodies against Ct have been demonstrated in animal models<sup>13</sup> and *in vitro*<sup>14</sup>. Paradoxically in a longitudinal

<sup>1</sup>Clinical Research Department, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, WC1E 7HT, United Kingdom. <sup>2</sup>ImmPORT Therapeutics, Inc./Antigen Discovery Inc., 1 Technology Dr., Suite E309, Irvine, CA, 92618, United States. <sup>3</sup>Disease Control and Elimination Theme, Medical Research Council The Gambia Unit, Fajara, Banjul, The Gambia. <sup>4</sup>Programa Nacional de Saúde de Visão, Ministério de Saúde Pública, Bissau, Guinea-Bissau. <sup>5</sup>Department of Pathogen Molecular Biology, London School of Hygiene and Tropical Medicine, London, WC1E 7HT, United Kingdom. <sup>6</sup>Pathogen Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom. <sup>7</sup>Institute of Structural and Molecular Biology, Birkbeck & University College London, Malet Street, London, WC1E 7HX, United Kingdom. <sup>8</sup>Department of Pathology and Laboratory Medicine, Medical Sciences I, Room D440, University of California, Irvine, CA, 92697-4800, United States. Correspondence and requests for materials should be addressed to H.P. (email: [harry.pickering@lshtm.ac.uk](mailto:harry.pickering@lshtm.ac.uk))

	Resistant	Susceptible	p-value
<b>Number</b>	60	30	NA
<b>Age in years (95% CI)</b>	9.00 (2.00–11.53)	8.00 (1.73–12.00)	0.413
<b>Female (n [%])</b>	24 (40.00)	13 (43.33)	0.762
<b>Village (n [%])</b>			<0.001
—1	7 (11.67)	19 (63.33)	
—2	2 (3.33)	0 (0.00)	
—4	5 (8.33)	1 (3.33)	
—6	11 (18.33)	3 (10.00)	
—7	27 (45.00)	7 (23.33)	
—8	8 (13.33)	0 (0.00)	

**Table 1.** Age, gender and village membership in resistant and susceptible groups screened on the micro-array. Association with resistance or susceptibility to infection was determined using a generalised linear model. For age, numbers in parentheses are the 95% confidence intervals around the median. For gender and village, numbers in parentheses are the percentage. No individuals from villages 3 and 9 were screened on the micro-array.

study from The Gambia, higher IgG responses against the immunodominant major outer membrane protein (MOMP) were associated with higher rates of infection and higher titres increased the associated risk<sup>15</sup>. Ocular Ct infection clearly induces a strong humoral immune response, but its role in protection or pathology in humans *in vivo* remains unclear.

Screening whole-proteome arrays has become an effective method to describe the complete profile of antibody responses in infection and disease<sup>16</sup>. Studies utilising these proteome arrays have highlighted some common themes in humoral immune targets of human pathogens including functions in protein binding and catalytic activities, early or late expression in the developmental cycle and membrane localisation<sup>17</sup>.

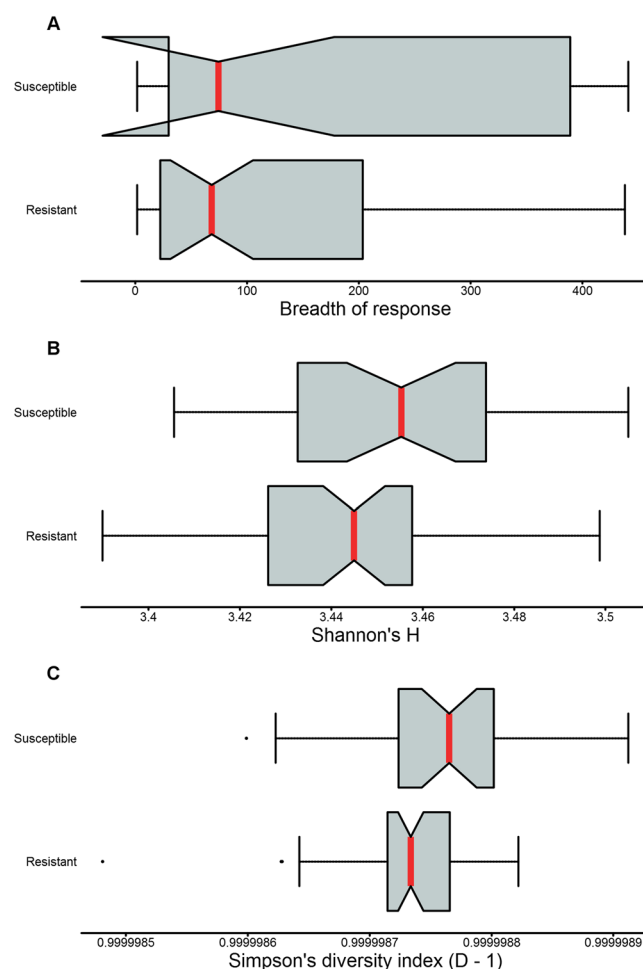
Genome-wide analyses of the type or mechanisms of selection have been conducted on a number of human pathogens and identified known and novel targets that were under immune selective pressures. Studies in many human pathogens, including *P. falciparum*<sup>18,19</sup>, *Helicobacter pylori*<sup>20</sup>, *Staphylococcus aureus*<sup>21</sup> and *Streptococcus pneumoniae*<sup>22</sup>, have highlighted cellular and humoral immune targets, cell surface proteins and known host-interactors as over-represented in genes under both balancing (selection and maintenance of multiple alleles) and positive selection (selection of advantageous alleles).

Until recently, the number of sequenced Ct genomes has been low and they have been derived from specimens that were collected from ecologically disparate sites over a timespan >50 years (Harris *et al.*<sup>23</sup>). A study by Thomson *et al.*<sup>24</sup> utilizing 3 genomes, more recent studies by Joseph *et al.*<sup>25,26</sup> utilising 12 and 32 genomes respectively followed by Borges *et al.*<sup>27</sup> (n = 59 genomes), identified positive selection in known host-interactors that were either surface-exposed or secreted into the host cytosol. Most recently, Hadfield *et al.* analysed 563 Ct genomes<sup>28</sup>. Similar to the prior studies, they focussed on global diversity of the species and its evolutionary history, rather than within-population dynamics. In trachoma, previous studies have focused on the population genetics of *ompA*, which encodes the immunodominant MOMP. Two populations of Ct sequences from trachoma-endemic communities in The Gambia<sup>29</sup> and Tanzania<sup>30</sup> suggested *ompA* was under purifying (selection against deleterious alleles) and positive selection; similar variation in selection pressure has been found in urogenital Ct sequences<sup>31</sup>. A lack of balancing selection in this immunodominant antigen, which is a target of neutralising antibodies, is in contrast to other pathogens and highlights the need for further population-based studies of Ct-genomes under selection to better understand the interactions between Ct and the host immune system.

We used sera, collected from Gambian children at the baseline point of a six-month longitudinal cohort, to screen a protein microarray of 894 genomic ORFs from serovar D Ct<sup>32</sup>. The complete profile of responses for each sample was determined and used to investigate the differential recognition of individual antigens and estimate the diversity and evenness of the antibody response associated with the frequency and duration of Ct infection. A population of 126 complete genome sequences of ocular Ct samples obtained from discrete communities of four Bijagos Islands (Guinea-Bissau) collected in a single survey<sup>33,34</sup> was used in tests of population genetic selection. Genome-wide evidence of selection and genome-wide screening of the antibody response in the context of susceptibility to ocular infection observed over a 6-month period were overlaid. This enabled us to identify new targets of humoral immunity and we uncovered two complementary immune evasion tactics that may support Ct survival and promote recurrent infection.

## Results

**Immunity defined by susceptibility to infection.** After normalisation and filtering to remove infrequently recognised antigens, responses of 90 individuals covering 441 antigens were included in the analysis. Individuals were divided into those resistant or susceptible based on observed median duration of infections over the six-month study period (Supplementary Figure 1). The study-wide median duration of infection was 2 weeks (or one visit). Those with no infections or short duration infections were combined ( $\leq 2$  weeks; resistant) and compared to those with long duration infections ( $> 2$  weeks; susceptible). The demographic similarity of resistant and susceptible individuals indicated that history of ocular Ct exposure was similar (Table 1).



**Figure 1.** Breadth and diversity of responses in resistant and susceptible individuals. Notched boxplots of breadth/diversity of responses (x-axis) in resistant and susceptible individuals (y-axis). (A) Breadth measured as the number positive responses within individuals ( $p = 0.088$ ). Diversity measured using (B) Shannon's diversity index ( $p = 0.024$ ) and (C) Simpson's diversity index ( $p = 0.080$ ). Median (red lines) and notches were calculated as the median  $\pm 1.57 \times \text{IQR}/\sqrt{n}$ , where IQR is the interquartile range and  $n$  is the number of samples. The IQR times 1.5 was added to the 75th percentile and subtracted from the 25th percentile to determine the whiskers. Dots are outliers.

**Diversity of the antibody response and infection susceptibility.** The complete profile of antibody responses in resistant and susceptible individuals was used to investigate differences in diversity or evenness of anti-Ct responses. Breadth, defined as the number of antigens recognised by an individual, was higher in susceptible individuals but this was not significant ( $p = 0.088$ ) (Fig. 1A). Diversity indices, which incorporate breadth and the relative strength of responses against an antigen, were higher in susceptible individuals. This reached significance for Shannon's ( $p = 0.024$ ) but not Simpson's diversity indices ( $p = 0.080$ ) (Fig. 1B and C). Together these indicate broader, less focussed antibody responses in susceptible individuals.

**Individual antibody responses were associated with increased susceptibility.** Association between responses to individual antigens and infection frequency and duration was determined using a generalised linear model adjusting for other major risk factors (age, gender and village of residence). Forty-two antigens were identified as targets of differential antibody responses between resistant and susceptible individuals ( $p \leq 0.05$ ) (Table 2). Higher responses to each of these antigens was associated with susceptibility to infection.

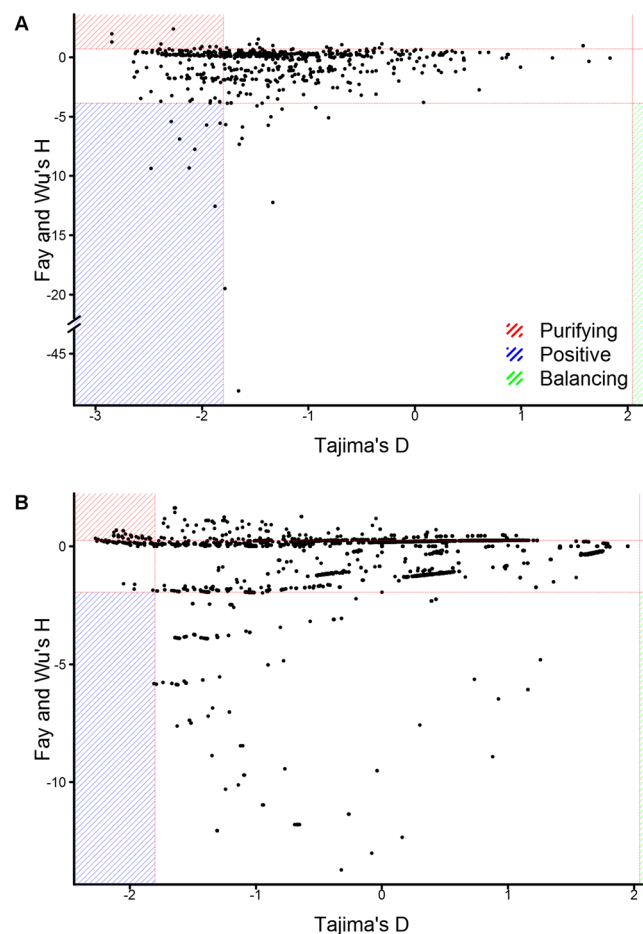
To examine when and how these antigens might be targeted by the host immune response during the Ct developmental cycle, their expression stage from Belland *et al.*<sup>35</sup> and predicted localisation from three computational tools (Ictree3, Cello and pSORTB) were compared between the 42 differentially recognised antigens and the 441 antigens (Supplementary Table 1). A  $\chi^2$  test was used to quantify over-representation or under-representation of expression stage or predicted localisations in the differentially recognised antigens, compared to all 441 antigens. Very early (1 hour post infection [HPI]) and very late (24–36 HPI) expressed genes were significantly over-represented ( $p = 0.007$ ). Secreted, inner membrane and periplasmic proteins were weakly over-represented in these antigens ( $p = 0.056$ ). Experimentally defined localisation of these antigens showed mixed agreement with software predictions, a number of those known to be secreted or reside in the outer membrane were incorrectly classified as remaining inside the inclusion. There was less individually-determined data available for expression

Antigen	p-value	p*	t	SE(t)	OR*	95% CI*	AUC
CT545	0.001	< 0.001	8.38	2.63	9.05	2.57–39.71	0.78
CT118	0.003	0.002	6.40	2.18	8.18	2.19–37.48	0.78
CT123	0.004	0.001	6.43	2.21	8.7	2.23–43.16	0.76
CT541	0.006	0.004	6.09	2.21	6.74	1.85–28.71	0.78
CT664	0.009	0.008	5.76	2.20	6.36	1.71–28.27	0.80
CT119	0.009	0.007	5.19	1.98	4.35	1.51–14.00	0.76
CT381	0.010	0.009	5.03	1.96	4.56	1.48–15.49	0.76
CT584	0.012	0.010	6.90	2.75	6.3	1.62–29.64	0.76
CT284	0.012	0.009	6.34	2.53	4.83	1.51–18.22	0.74
CT502	0.014	0.012	6.46	2.62	5	1.50–19.93	0.80
CT051	0.017	0.015	5.06	2.11	5.44	1.47–24.26	0.74
CT592	0.019	0.018	4.14	1.77	3.6	1.28–11.23	0.75
CT073	0.020	0.017	6.01	2.58	6.91	1.47–39.52	0.77
CT078	0.020	0.018	6.07	2.61	4.88	1.38–20.58	0.74
CT023	0.022	0.019	6.60	2.88	7.61	1.56–51.76	0.77
CT223	0.022	0.020	5.93	2.59	4.11	1.31–15.14	0.76
CT795	0.024	0.018	3.91	1.73	3.67	1.23–11.93	0.76
CT875	0.028	0.030	5.14	2.34	4.21	1.22–16.47	0.73
CT181	0.030	0.024	4.32	1.99	4.05	1.20–15.42	0.75
CT694	0.031	0.031	3.91	1.81	3.68	1.17–12.78	0.74
CT021	0.034	0.033	5.45	2.57	4.72	1.19–21.77	0.77
CT813	0.035	0.033	3.50	1.66	3.13	1.12–9.57	0.75
CT841	0.036	0.035	3.77	1.80	3.15	1.12–9.78	0.74
CT017	0.036	0.039	3.77	1.80	4.37	1.16–18.92	0.75
CT142	0.036	0.037	4.24	2.03	3.41	1.12–11.48	0.74
CT764	0.037	0.034	5.74	2.75	4.5	1.18–20.92	0.78
CT728	0.039	0.040	4.86	2.35	3.97	1.14–16.23	0.78
CT228	0.039	0.037	5.54	2.69	4.34	1.16–19.64	0.74
CT703	0.040	0.042	4.82	2.34	5.76	1.20–34.90	0.74
CT494	0.040	0.037	5.63	2.74	5.62	1.19–32.96	0.76
CT106	0.041	0.039	5.87	2.87	3.45	1.11–12.30	0.76
CT097	0.041	0.040	5.50	2.70	3.6	1.10–13.26	0.75
CT316	0.043	0.049	6.00	2.96	3.66	1.10–14.02	0.75
CT579	0.044	0.041	3.73	1.85	3.55	1.09–13.26	0.77
CT168	0.044	0.044	6.32	3.14	5.37	1.19–32.48	0.76
CT089	0.044	0.045	4.54	2.26	3.34	1.07–11.61	0.75
CT237	0.045	0.043	5.79	2.88	4.53	1.13–22.39	0.75
CT806	0.046	0.046	4.89	2.45	3.49	1.08–12.84	0.76
CT668	0.048	0.049	5.68	2.87	3.45	1.07–12.91	0.77
CT642	0.048	0.047	5.80	2.93	3.39	1.08–12.52	0.76
CT695	0.049	0.053	3.73	1.90	3.44	1.06–12.84	0.75
CT570	0.050	0.049	5.92	3.02	4.29	1.08–20.48	0.76

**Table 2.** Forty-two differentially recognised antigens between resistant and susceptible individuals. Univariate associations were determined using a generalised linear model. Variables were resampled 10,000 times and remodelled to determine permuted p-values (P\*). Odds ratios (OR) and confidence intervals (CI) were calculated for an increase of half the range per antigen, rather than one unit. The predictive value of the univariate generalised linear models was calculated and presented as area under the curve (AUC).

stage, although some antigens predicted to peak at 40 HPI have been shown since to peak 1–6 HPI. These disagreements, in classification of expression stage and localisation, strengthen the over-representation of antigens expressed early or late and localised to interact directly with the host.

**Genome-wide evidence of purifying and positive selection.** In general, pathogen immunogenic proteins are under natural selection, due to their impact on pathogen survival and transmission. To validate identified immune targets and identify further targets, sequence data from 126 ocular Ct samples from the Bjiagos Islands, Guinea-Bissau was examined for evidence of departure from neutral selection. These samples were collected and sequenced as described elsewhere<sup>33,34</sup>, details in Supplementary Methods. Currently, no Ct whole-genome sequences are available from the villages included in this study. To determine the relevance of evidence of selection in samples from the Bijagos Islands, they were compared with five historical isolates from



**Figure 2.** Correlation between Tajima's D and Fay and Wu's H. Genome-wide correlation of D and H values at (A) gene-level and (B) epitope-level (sliding windows of 42 bp). Values significantly different from zero are indicated for each measure (dashed red lines). Genes with evidence of positive (blue), purifying (red) and balancing selection (green) are highlighted.

The Gambia (Supplementary Figure 2). The Bijagos Islands samples were separated from each other by 1–1119 SNPs, the Gambian samples were separated by 161–1704 SNPs and the two populations of samples were separated by 487–1019 SNPs. As expected, based on previous Ct sequences, this suggests the geographically-distinct populations are close genetic relations, supporting the use of the Bijagos Islands samples in predicting genes under selection within those circulating in The Gambia. Therefore we tested for evidence of departure from neutrality using a number of tests including Tajima's D, Fay and Wu's H, and the integrated haplotype score.

Tajima's D can distinguish between directional selection (positive or purifying selection [ $D \leq -1.8$ ]) and balancing selection ( $D \geq 2.044$ ), by comparing the levels of low and medium frequency alleles<sup>36</sup>. Fay and Wu's H can distinguish between balancing/positive selection and balancing/purifying selection, by comparing the levels of low and high frequency alleles. Balancing ( $D \geq 2.044$  and  $0.72 \leq H \leq -3.85$ ), positive ( $D \leq -1.8$  and  $H \leq -3.85$ ) and purifying selection ( $D \leq -1.8$  and  $H \geq 0.72$ ) can be differentiated, by combining D- and H-values.

Nineteen genes with evidence of selection by Tajima's D were supported by Fay and Wu's H. Ten of these genes had evidence of positive selection and 9 genes had evidence of purifying selection (Fig. 2A and Table 3 ['Windows under selection' = 0]). D- and H-values were then determined using a 42 base pair sliding window analysis, this window equates to the most frequent length of an antibody epitope (16 amino acids). Seventy-six windows across 12 genes had evidence of positive selection and 61 windows across 8 genes had evidence of purifying selection (Fig. 2B and Table 3 ['Windows under selection' > 0]). Evidence of natural selection acting on/within these genes suggests they may influence Ct survival and transmission.

### Integrated haplotype scores identify three genomic regions under positive selection.

Identification of positive selection by Tajima's D and Fay and Wu's H is most powerful when an allele is close to fixation (only one allele at a given site). A genome-wide scan was performed to calculate integrated haplotype scores (iHS) for SNPs, to identify genes and regions under positive selection that have not yet reached fixation.

The median iHS score was 0.66 (95% CI 0.03–2.18), 20 SNPs in the top 1% of the genome-wide distribution of iHS (Fig. 3). The top 1% of SNPs highlighted three loci, which showed evidence of recent positive selection;

ID	Name	Number of SNPs	Fay and Wu's H	Tajima's D	Windows under selection	Type of Selection
CT033	<i>recD_1</i>	15	−5.41	−2.29	0	Positive
CT046	<i>hctB</i>	33	1.00	−1.33	2	Purifying
CT082		12	−9.32	−2.12	1	Positive
CT105		45	2.40	−2.27	13	Purifying
CT147		32	−12.23	−1.33	24	Positive
CT159		19	−3.23	−2.28	4	Positive
CT223	<i>IPAM</i>	15	0.91	−1.87	0	Purifying
CT249		5	−5.67	−1.78	1	Positive
CT288		12	−12.55	−1.88	0	Positive
CT359		8	−5.55	−1.83	1	Positive
CT386		7	−7.75	−2.07	0	Positive
CT394	<i>hrcA</i>	15	0.73	−2.01	0	Purifying
CT414	<i>pmpC</i>	41	−2.88	−1.52	16	Purifying
CT442	<i>crpA</i>	28	−0.86	−2.52	17	Purifying
CT456	<i>tarP</i>	32	−6.84	−1.62	8	Positive
CT539	<i>trxA</i>	4	−3.87	−1.76	2	Positive
CT621		16	0.90	−1.93	0	Purifying
CT622		47	−48.13	−1.66	4	Positive
CT624	<i>mviN</i>	17	0.91	−2.00	0	Purifying
CT636	<i>greA</i>	15	0.73	−2.07	0	Purifying
CT641	<i>ygeD</i>	42	1.13	−2.38	0	Purifying
CT651		51	0.32	−2.63	0	Purifying
CT674	<i>yscC</i>	51	1.30	−2.85	0	Purifying
CT681	<i>ompA</i>	25	−0.46	−2.20	13	Purifying
CT686	<i>sufD</i>	20	−19.50	−1.78	21	Positive
CT688	<i>parB</i>	7	−5.71	−1.96	3	Positive
CT694		14	−6.89	−2.21	6	Positive
CT868	<i>Dub1</i>	19	−9.36	−2.48	0	Positive
CT872	<i>pmpH</i>	68	1.99	−2.85	0	Purifying

**Table 3.** Genes under selection identified by Tajima's D and Fay and Wu's H. Number of SNPs per gene and the gene-level Tajima's D and Fay and Wu's H values are indicated. The number of sliding windows with significant evidence of selection by both Fay and Wu's H and Tajima's D are indicated ('Windows under selection'). Type of selection is indicated.

CT048-CT074, CT154-CT155 and CT456-CT625. These loci include a region covering *tarP* and *pmp* family members and a region within the Ct plasticity zone.

**Ct genes important for Ct survival and pathogenesis were the focus of natural selection.** In this population of ocular Ct samples, 48 genes were identified with evidence of selection by either a combination of Tajima's D and Fay and Wu's H, at the gene or epitope level, or by iHS. Expression stage of these genes and localisation of the translated proteins was examined to discover common patterns in genes with evidence of selection (Supplementary Table 2). Secreted and outer membrane proteins were significantly over-represented in these targets ( $p = 0.004$ ), as were genes with peak expression levels very early or very late in the developmental cycle ( $p = 0.0005$ ). Expression at these pivotal stages of infection and extra-inclusion localisations suggests these genes are important factors in Ct survival and pathogenesis. Similar to the 42 antigens identified in this study, comparison of the array-determined expression stage and predicted localisation to more recent experiments supported these findings.

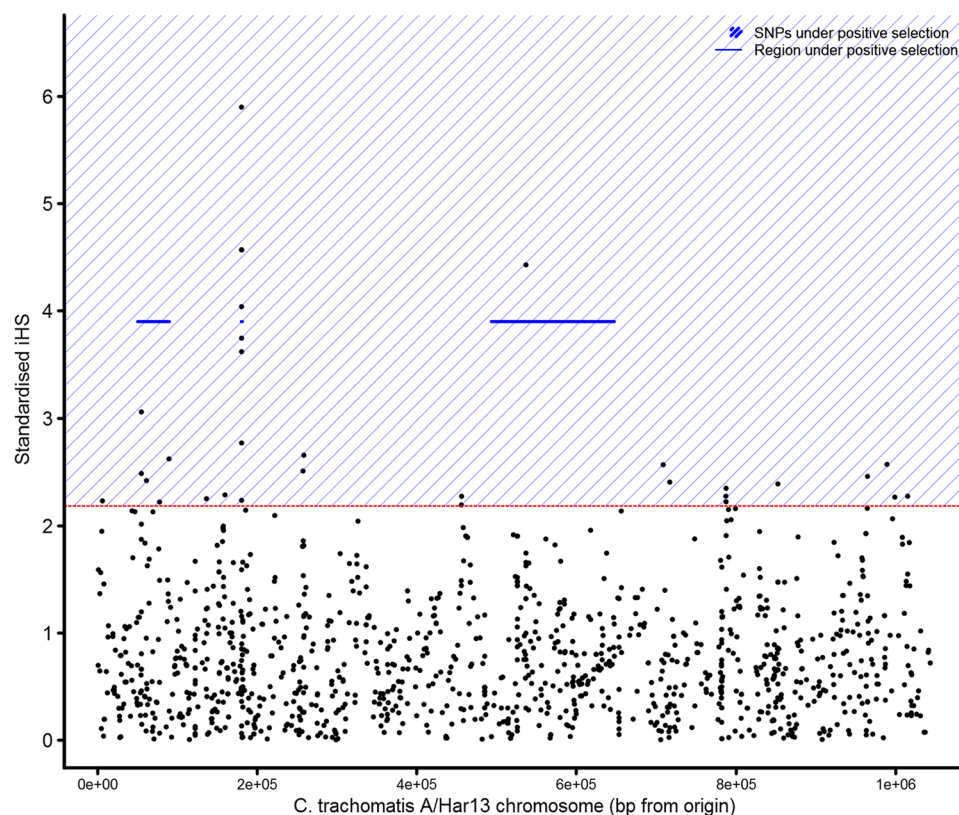
**Variable evidence of selection acting on genes associated with infection frequency and duration.** Evidence of selection is a common marker of pathogen immunogenic proteins due to their interactions with the host and impact on pathogen survival and transmission. Therefore, we examined the 42 antigens associated with susceptibility to infection for evidence of selection, as one means of validation as important immune targets (Fig. 4 and Supplementary Table 3).

Four of 42 targets associated with susceptibility to infection had gene-level evidence of selection by Tajima's D, supported by Fay and Wu's H. CT694 and CT695, had evidence of positive selection at the epitope level supported by D and H values. CT545 and CT806 had evidence of purifying selection at the epitope level.

Ten genes were within regions under positive selection by iHS. One target, susceptibility-associated CT228, contained a SNP under positive selection by iHS.

The majority of the antigens associated with susceptibility had no evidence of selection in this population, suggesting they are evolving under neutral selection.





**Figure 3.** Evidence of positive selection using the integrated haplotype score. Genes in the top 1% of values (dashed red line) had the strongest evidence of positive selection. Regions (blue lines) and individual SNPs (blue shading) under positive selection are indicated.

## Discussion

There is considerable evidence from animal models suggesting antibody responses are necessary for long-term protection and immunity from chlamydial infection<sup>37–39</sup>. In mice, the breadth of the antibody response is higher in strains that are more susceptible to chlamydial infection and associated pathology<sup>40–42</sup>. Furthermore, in non-human primates, partial immunity to ocular infection was consistent with development of a focussed antibody recall response<sup>43</sup>.

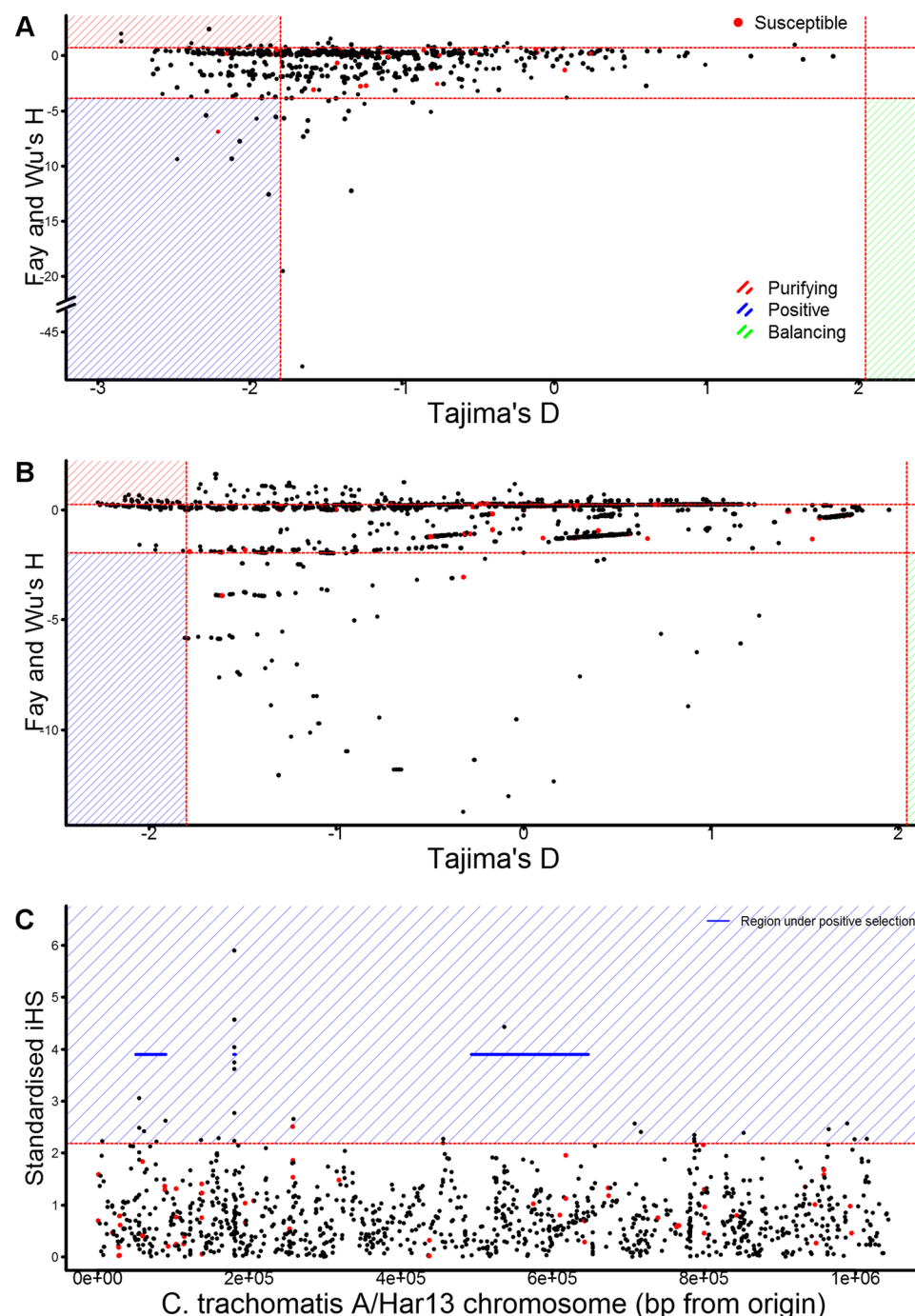
Previous work on the Gambian six-month longitudinal cohort utilised in this study focussed on systemic and local cell-mediated immune responses<sup>44, 45</sup>. To gain a more complete picture of the immune responses underlying differential outcomes, we screened serum samples against a Ct protein microarray to examine the relationship between individual serological immune responses and the acquisition and resolution of ocular Ct infection. Antibody responses were more focused in children who were able to resolve infection, while heightened responses to 42 antigens were associated with susceptibility to infection and longer durations of infection.

Tests for selection identified a number of genes and regions of the Ct genome under purifying and positive selection, using 126 ocular samples collected from the Bijagos Islands (Guinea-Bissau). These regions were focused in immunogenic proteins and those that interact directly with the host. This result was further strengthened by using experimentally derived expression stage and localisation of individually investigated genes, as opposed to array-based profiling of expression and bioinformatics prediction of localisation. Evidence for selection within the potential immune targets identified from the proteome array analysis was variable, 5/42 targets had significant evidence of selection.

Genes important in host-pathogen interactions are under natural selection. Most frequently, selection is observed in pathogen immune targets, where balancing and/or positive selection can aid immune evasion<sup>46</sup>. We found evidence of selection in Ct genes that code for known or putative immune targets or virulence factors, similar to previously described Ct genomics studies<sup>24–27</sup>. Genes coding for proteins involved in cell entry, cell exit and intracellular interactions via the inclusion membrane encompassed the majority of genes under selection. Additionally, 17 of 48 genes with evidence of selection have no known function, suggesting these genes may be important for Ct survival and transmission. Genome-wide scans for evidence of selection therefore have the ability to highlight important, previously uncharacterized genes.

Seventeen of the genes under selection are known to be immunogenic; this supports immune-recognition as a key driving factor of selection within this population of ocular Ct samples. However, there was limited evidence of balancing selection, a common mechanism of immune evasion employed by human pathogens<sup>18, 21, 47, 48</sup>. This method of immune evasion relies on cyclical presentation of alternative forms of immunodominant, primarily surface-exposed, antigens to the host immune system to avoid recognition and clearance.





**Figure 4.** Evidence of selection in antibody targets associated with susceptibility to infection. Evidence of selection was determined at the gene-level (A) and the epitope-level (B) by Tajima's D and Fay and Wu's H. Evidence of positive selection on SNPs and larger genomic regions was independently determined by iHS (C). Association with susceptibility to infection is indicated (red). Thresholds for genes considered under selection, values significantly different from zero, is indicated for each measure (dashed red lines). Genes or SNPs (C) with evidence of positive (blue), purifying (red) and balancing selection (green) are highlighted.

Typically, in trachoma endemic communities where immunity develops slowly, Ct is able to reinfect individuals within the same households or village. The lack of protective antibodies, combined with the absence of balancing selection, suggests that Ct employs a different strategy for immune evasion. The results suggest two potential routes for immune evasion, a) blocking and/or invasion-enhancing antibodies against surface antigens or b) masking of protective humoral immune responses through heightened responses against a large number of immunogenic but non-protective antigens.

Blocking antibodies are induced by surface antigens of *P. falciparum*<sup>49</sup>, *Candida albicans* and *Neisseria gonorrhoeae* and can inhibit protective responses<sup>50,51</sup>. Antibody-dependent enhancement is a well-described process

in viral infections<sup>52,53</sup>, involving cross-linking of host-cell surface receptors by virus-antibody complexes leading to enhanced infectivity. Similar observations in *P. vivax*<sup>54</sup> suggest this mechanism may be more widely utilized by pathogens.

Through Ct-protein array screening of serum from children resistant or susceptible to frequent and/or prolonged Ct infection, we identified three susceptibility associated surface antigens CT017 (Ctd1), CT541 (MIP) and CT579. Antibodies generated against non-protective surface antigens provide a survival advantage for Ct and such mutations would be expected to be under purifying selection. There was no clear evidence of selection in these antigens. However, three of four outer membrane proteins (*ompA* [MOMP], *pmpC* and *pmpH*) with strong evidence of selection were under purifying selection, suggesting these antigens are targets of blocking and/or invasion-enhancing antibodies. In support of this, MOMP can induce blocking<sup>55</sup> and invasion-enhancing antibodies<sup>14,56</sup> *in vitro* and in mouse models.

The majority of the 42 susceptibility associated antigens were not localised to the EB surface, therefore their role in Ct infection and disease cannot be explained by a hypothesis that solely involves non-protective surface antigens. The breadth of the antibody response shows that a high number of Ct antigens are accessible to the host immune system, likely through release of Ct proteins as a result of inclusion/host cell lysis or recently described chlamydial extrusions<sup>57,58</sup>. Antibody responses against the majority of Ct proteins are not protective, leading to a hypothesis that heightened responses to a large number of immunogenic but not protective antigens masks protective humoral immune responses.

A diverse antibody profile would provide an advantage for the bacteria therefore targeted genes would be expected to be under purifying selection, however the number of antigens targeted and their inherently random nature implies no one gene would show strong evidence of selection. In support of evasion by decoy, 37/42 of the antigens identified as susceptibility-associated had no strong evidence of selection. Of the five under selection, two had evidence of purifying selection (CT545 and CT806). The remaining three were under positive selection (CT228, CT694 and CT695), however they are known virulence factors<sup>59–63</sup>, therefore immune evasion may not be the only driving force of selection.

Ideally, samples from the Gambian villages utilised in this study would have been screened for evidence of selection, however whole-genome sequences are currently unavailable. The phylogeny of the Bijagos Islands samples and the previously published Gambian samples, suggests they are closely related. This relatedness and the similar community-level endemicity of active trachoma (Gambian cohort 21.5% and Bijagos Islands 22% [one to nine year olds]) and ocular Ct infection (Gambian cohort 9.9% and Bijagos Islands 25% [one to nine year olds]), support the assertion that the genes under selection may be shared between these geographically-separated Ct populations. Additionally, evidence of diversity within the Bijagos Islands samples and historical Gambian samples, suggests observed differences in infection frequency and duration between villages are not due to the presence of a single dominant Ct clone.

These data demonstrate the value of genomic information in the identification of immune targets and virulence factors, particularly in combination with proteome-wide antibody responses. Less focussed antibody responses in susceptible individuals and evidence of purifying selection in Ct surface antigens, support an evasion strategy in which Ct presents a number of non-protective, irrelevant antigens to the immune system to block or misdirect protective responses. These results will allow targeted development of vaccine candidates and may explain the observed success of single and multi-antigen vaccines against Ct, as they would promote a focussed, protective humoral immune response. In summary, our data support the importance of antibody responses in human immunity to Ct infection.

## Methods

**Ethics statement.** The Gambian longitudinal cohort was conducted in accordance with the Declaration of Helsinki. The study and its procedures were approved by the Gambia Government/Medical Research Council The Gambia Unit Joint Ethics Committee and by the Ethics Committee of the London School of Hygiene and Tropical Medicine (LSHTM). Verbal consent was obtained from community leaders. Written informed consent was obtained from all study participants' guardians on their behalf. A signature or thumbprint is considered an appropriate record of consent in this setting by the above ethical bodies.

The Bissau-Guinean cross-sectional survey was conducted in accordance with the Declaration of Helsinki. The study and its procedures were approved by the Comitê Nacional de Ética e Saúde (Guinea-Bissau) and the LSHTM Ethics Committee. Verbal consent was obtained from community leaders. Written informed consent was obtained from all study participants or their guardians on their behalf if participants were children. A signature or thumbprint is considered an appropriate record of consent in this setting by the above ethical bodies.

**Clinical cohort study and participants.** A six-month longitudinal study was previously conducted and recruited 345 children, aged four to fifteen years old, from nine villages in The Gambia<sup>11</sup>. Villages were selected following initial trachoma rapid assessment screening that found clinical signs of active trachoma in 20% of school age children. At baseline and at each fortnightly visit for six-months, children were examined for clinical signs of active trachoma. Trachoma was graded according to the simplified WHO grading system<sup>64</sup>. Two conjunctival swabs were collected in duplicate to test for Ct infection and tear fluid was collected for mucosal cytokine and antibody assays. Venous blood samples were collected at baseline and cessation of the study, 186 samples were collected. Villages 3 and 9 declined to give venous blood samples and could not be included in this study.

**Chlamydia trachomatis antigen microarrays.** Ninety sera from baseline and 33 sera from the end of the study were screened against a previously published protein microarray covering 894 genomic open reading frames (ORFs) from serovar D Ct<sup>32</sup>. Briefly, 894 ORFs were cloned into the pXT7 expression vector, proteins were expressed and printed on glass slides. After blocking, the arrays were probed with diluted sera, bound antibody was detected using a biotin-conjugated anti-human antibody followed by a streptavidin-conjugated secondary antibody.

**Proteome microarray normalisation, filtering and clustering.** The raw signal intensity data (Supplementary Table 4) from the protein micro-array was transformed by inverse hyperbolic sine transformation and normalised by mean-centring. To filter out infrequently recognised antigens, post-normalisation, the global median of the data was calculated and individual antigens whose median was lower than the global median were excluded from further analysis.

Different methods to identify positive-negative breakpoints in the distribution of the data were tested. Silhouette width was used to quantify the best method. Cluster separation per antigen was derived from the average silhouette width. To classify positive responses, two clusters were identified using the method that had the highest average silhouette width for each antigen.

**Diversity metrics.** Breadth of response was defined as the number of antigens to which each individual had a positive response. Diversity was calculated using Shannon's entropy (H) and Simpson's diversity index (D).

**Statistical analyses.** Intensity of responses was compared using a generalised linear model (glm) and the number of positive responses using Fisher's exact test. For the glm, 10,000 permutations of the outcome variable were performed to generate an adjusted p-value. Adjusted odds ratios (OD\*) and adjusted confidence intervals (CI\*) were calculated as the unadjusted OR/CI exponentiated by half the range for a given antigen.

**Chlamydia trachomatis population genetic metrics.** As part of a cross-sectional population-based survey in trachoma-endemic communities on the Bijagós Archipelago of Guinea-Bissau, upper tarsal conjunctival swabs were taken<sup>33</sup>. Whole-genome sequence data (Supplementary Table 5) was obtained from 126 Ct-positive swabs using SureSelect enrichment and Illumina paired-end DNA sequencing technology and assembled using the A/Har13 as reference genome. Aligned multi-fasta files for each gene were used as input for Variscan-2.0.3<sup>65</sup> to calculate Tajima's D and Fay and Wu's H. Integrated haplotype scores (iHS) were calculated using the rehh package in R. Missing base-calls were imputed using a genetic-distance based method.

## References

1. WHO. *Trachoma: Fact Sheet*, <http://www.who.int/mediacentre/factsheets/fs382/en/> (2016).
2. Bailey, R. L. *et al.* The duration of human ocular *Chlamydia trachomatis* infection is age dependent. *Epidemiology and Infection* **123**, 479–486 (1999).
3. Taylor, H. R. *et al.* The natural history of endemic trachoma: a longitudinal study. *The American journal of tropical medicine and hygiene* **46**, 552–559 (1992).
4. Treharne, J. D. The community epidemiology of trachoma. *Reviews in Infectious Diseases* **7**, 760–764 (1985).
5. Rasmussen, S. J. *et al.* Secretion of proinflammatory cytokines by epithelial cells in response to *Chlamydia* infection suggests a central role for epithelial cells in chlamydial pathogenesis. *The Journal of Clinical Investigation* **99**, 77–87 (1997).
6. el-Asrar, A. M. *et al.* Immunopathology of trachomatous conjunctivitis. *British Journal of Ophthalmology* **73**, 276–282 (1989).
7. Natividad, A. *et al.* Human conjunctival transcriptome analysis reveals the prominence of innate defense in *Chlamydia trachomatis* infection. *Infection and immunity* **78**, 4895–4911 (2010).
8. Bailey, R. L., Holland, M. J., Whittle, H. C. & Mabey, D. C. W. Subjects Recovering from Human Ocular Chlamydial Infection Have Enhanced Lymphoproliferative Responses to Chlamydial Antigens Compared with Those of Persistently Diseased Controls. *Infection and immunity* **63**, 389–392 (1995).
9. Holland, M. J. *et al.* Synthetic peptides based on *Chlamydia trachomatis* antigens identify cytotoxic T lymphocyte responses in subjects from a trachoma-endemic population. *Clinical and experimental immunology* **107**, 44–49 (1997).
10. Brunham, R. C. & Rey-Ladino, J. Immunology of Chlamydia infection: implications for a *Chlamydia trachomatis* vaccine. *Nature Reviews Immunology* **5**, 149–161 (2005).
11. Faal, N. *et al.* Conjunctival FOXP3 expression in trachoma: do regulatory T cells have a role in human ocular *Chlamydia trachomatis* infection? *PLoS medicine* **3**, e266 (2006).
12. Collier, L. H., Sowa, J. & Sowa, S. The serum and conjunctival antibody response to trachoma in Gambian children. *Journal of Hygiene (Lond)* **70**, 727–740 (1972).
13. Barenfanger, J. & MacDonald, A. B. The role of immunoglobulin in the neutralization of trachoma infectivity. *Journal of immunology* **113**, 1607–1617 (1974).
14. Peterson, E. M., Cheng, X., Pal, S. & de la Maza, L. M. Effects of Antibody Isotype and Host Cell Type on *In Vitro* Neutralization of *Chlamydia trachomatis*. *Infection and immunity* **61**, 498–503 (1993).
15. Bailey, R. L. *et al.* The influence of local antichlamydial antibody on the acquisition and persistence of human ocular chlamydial infection: IgG antibodies are not protective. *Epidemiology and Infection* **111**, 315–324 (1993).
16. Davies, D. H. *et al.* Profiling the humoral immune response to infection by using proteome microarrays: high-throughput vaccine and diagnostic antigen discovery. *Proceedings of the National Academy of Sciences United States of America* **102**, 547–552 (2005).
17. Liang, L. & Felgner, P. L. A systems biology approach for diagnostic and vaccine antigen discovery in tropical infectious diseases. *Current opinion in infectious diseases* **28**, 438–445 (2015).
18. Amambua-Ngwa, A. *et al.* Population Genomic Scan for Candidate Signatures of Balancing Selection to Guide Antigen Characterization in Malaria Parasites. *PLoS Genetics* **8**, e1002992 (2012).
19. Yuan, L. *et al.* *Plasmodium falciparum* populations from northeastern Myanmar display high levels of genetic diversity at multiple antigenic loci. *Acta tropica* **125**, 53–59 (2013).
20. Yahara, K. *et al.* Genome-wide survey of codons under diversifying selection in a highly recombining bacterial species, *Helicobacter pylori*. *DNA Research* **23**, 135–143 (2016).
21. Thomas, J. C., Godfrey, P. A., Feldgarden, M. & Robinson, D. A. Candidate targets of balancing selection in the genome of *Staphylococcus aureus*. *Molecular biology and evolution* **29**, 1175–1186 (2012).
22. Li, Y. *et al.* Distinct effects on diversifying selection by two mechanisms of immunity against *Streptococcus pneumoniae*. *PLoS pathogens* **8**, e1002989 (2012).
23. Harris, S. R. *et al.* Whole-genome analysis of diverse *Chlamydia trachomatis* strains identifies phylogenetic relationships masked by current clinical typing. *Nature genetics* **44**, 413–419 (2012).
24. Thomson, N. R. *et al.* *Chlamydia trachomatis*: genome sequence analysis of lymphogranuloma venereum isolates. *Genome research* **18**, 161–171 (2008).
25. Joseph, S. J. *et al.* Interplay of recombination and selection in the genomes of *Chlamydia trachomatis*. *Biology direct* **6** (2011).
26. Joseph, S. J. *et al.* Population genomics of *Chlamydia trachomatis*: insights on drift, selection, recombination, and population structure. *Molecular biology and evolution* **29**, 3933–3946 (2012).

27. Borges, V. *et al.* Directional evolution of *Chlamydia trachomatis* towards niche-specific adaptation. *Journal of bacteriology* **194**, 6143–6153 (2012).
28. Hadfield, J. *et al.* Comprehensive global genome dynamics of *Chlamydia trachomatis* show ancient diversification followed by contemporary mixing and recent lineage expansion. *Genome research* (2017).
29. Andreasen, A. A. *et al.* *Chlamydia trachomatis* ompA variants in trachoma: what do they tell us? *PLoS neglected tropical diseases* **2**, e306 (2008).
30. Hsieh, Y. H., Bobo, L. D., Quinn, T. C. & West, S. K. Determinants of trachoma endemicity using *Chlamydia trachomatis* ompA DNA sequencing. *Microbes and Infection* **3**, 447–458 (2001).
31. Nunes, A. *et al.* Evolutionary dynamics of ompA, the gene encoding the *Chlamydia trachomatis* key antigen. *Journal of bacteriology* **191**, 7182–7192 (2009).
32. Patton, D. L. *et al.* Whole genome identification of *C. trachomatis* immunodominant antigens after genital tract infections and effect of antibiotic treatment of pigtailed macaques. *Journal of proteomics* **108**, 99–109 (2014).
33. Last, A. R. *et al.* Plasmid Copy Number and Disease Severity in Naturally Occurring Ocular *Chlamydia trachomatis* Infection. *Journal of clinical microbiology* **52**, 324–327 (2014).
34. Last, A. R. *et al.* Risk Factors for Active Trachoma and Ocular *Chlamydia trachomatis* Infection in Treatment-Naive Trachoma-Hyperendemic Communities of the Bijagos Archipelago. *PLoS neglected tropical diseases* **8**, e2900 (2014).
35. Belland, R. J. *et al.* Genomic transcriptional profiling of the developmental cycle of *Chlamydia trachomatis*. *Proceedings of the National Academy of Science of the United States of America* **100**, 8478–8483 (2003).
36. Tajima, F. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics* **123**, 585–595 (1989).
37. Farris, C. M., Morrison, S. G. & Morrison, R. P. CD4+ T cells and antibody are required for optimal major outer membrane protein vaccine-induced immunity to *Chlamydia muridarum* genital infection. *Infection and immunity* **78**, 4374–4383 (2010).
38. Morrison, S. G. & Morrison, R. P. A predominant role for antibody in acquired immunity to chlamydial genital tract reinfection. *Journal of immunology* **175**, 7536–7542 (2005).
39. Olsen, A. W. *et al.* Protection Against *Chlamydia trachomatis* Infection and Upper Genital Tract Pathological Changes by Vaccine-Promoted Neutralizing Antibodies Directed to the VD4 of the Major Outer Membrane Protein. *Journal of Infectious Diseases* **4**, 978–989 (2015).
40. de la Maza, L. M., Pal, S., Khamesipour, A. & Peterson, E. M. Intravaginal inoculation of mice with the *Chlamydia trachomatis* mouse pneumonitis biovar results in infertility. *Infection and immunity* **62**, 2094–2097 (1994).
41. Darville, T. *et al.* Mouse strain-dependent variation in the course and outcome of chlamydial genital tract infection is associated with differences in host response. *Infection and immunity* **65**, 3065–3073 (1997).
42. Teng, A. *et al.* Proteomic identification of immunodominant chlamydial antigens in a mouse model. *Journal of proteomics* **77**, 176–186 (2012).
43. Kari, L. *et al.* Antibody signature of spontaneous clearance of *Chlamydia trachomatis* ocular infection and partial resistance against re-challenge in a nonhuman primate trachoma model. *PLoS neglected tropical diseases* **7**, e2248 (2013).
44. Faal, N. *et al.* Temporal cytokine gene expression patterns in subjects with trachoma identify distinct conjunctival responses associated with infection. *Clinical and Experimental Immunology* **142**, 347–353 (2005).
45. Holland, M. J. *et al.* The frequency of *Chlamydia trachomatis* major outer membrane protein-specific CD8+ T lymphocytes in active trachoma is associated with current ocular infection. *Infection and immunity* **74**, 1565–1572 (2006).
46. Brunham, R. C., Plummer, F. A. & Stephens, R. S. Bacterial antigenic variation, host immune response, and pathogen-host coevolution. *Infection and immunity* **61**, 2273–2276 (1993).
47. Wang, I. N. *et al.* Genetic diversity of ospC in a local population of *Borrelia burgdorferi* sensu stricto. *Genetics* **151**, 15–30 (1999).
48. Tetteh, K. K. *et al.* Prospective identification of malaria parasite genes under balancing selection. *PLoS one* **4**, e5568 (2009).
49. Guevara Patino, J. A., Holder, A. A., McBride, J. S. & Blackman, M. J. Antibodies that inhibit malaria merozoite surface protein-1 processing and erythrocyte invasion are blocked by naturally acquired human antibodies. *Journal of Experimental Medicine* **186**, 1689–1699 (1997).
50. Bromuro, C. *et al.* Interplay between protective and inhibitory antibodies dictates the outcome of experimentally disseminated Candidiasis in recipients of a *Candida albicans* vaccine. *Infection and immunity* **70**, 5462–5470 (2002).
51. Rice, P. A., Vayo, H. E., Tam, M. R. & Blake, M. S. Immunoglobulin G antibodies directed against protein III block killing of serum-resistant *Neisseria gonorrhoeae* by immune serum. *Journal of Experimental Medicine* **164**, 1735–1748 (1986).
52. Ferguson, N., Anderson, R. & Gupta, S. The effect of antibody-dependent enhancement on the transmission dynamics and persistence of multiple-strain pathogens. *Proceedings of the National Academy of Science of the United States of America* **96**, 790–794 (1999).
53. Takada, A. & Kawaoka, Y. Antibody-dependent enhancement of viral infection: molecular mechanisms and *in vivo* implications. *Reviews in Medical Virology* **13**, 387–398 (2003).
54. Carter, R. *et al.* Immunity to sexual stages of malaria parasites. *Progress in Allergy* **41**, 193–214 (1988).
55. Crane, D. D. *et al.* *Chlamydia trachomatis* polymorphic membrane protein D is a species-common pan-neutralizing antigen. *Proceedings of the National Academy of Sciences United States of America* **103**, 1894–1899 (2006).
56. Su, H., Spangrude, G. J. & Caldwell, H. D. Expression of Fc gamma RIII on HeLa 229 cells: possible effect on *in vitro* neutralization of *Chlamydia trachomatis*. *Infection and immunity* **59**, 3811–3814 (1991).
57. Zuck, M. *et al.* Conservation of extrusion as an exit mechanism for *Chlamydia*. *Pathogens and disease* **74** (2016).
58. Shaw, J. H. *et al.* Comparison of Murine Cervicovaginal Infection by Chlamydial Strains: Identification of Extrusions Shed *In vivo*. *Frontiers in cellular and infection microbiology* **7** (2017).
59. Lutter, E. I., Barger, A. C., Nair, V. & Hackstadt, T. *Chlamydia trachomatis* inclusion membrane protein CT228 recruits elements of the myosin phosphatase pathway to regulate release mechanisms. *Cell reports* **3**, 1921–1931 (2013).
60. Hower, S., Wolf, K. & Fields, K. A. Evidence that CT694 is a novel *Chlamydia trachomatis* T3S substrate capable of functioning during invasion or early cycle development. *Molecular microbiology* **72**, 1423–1437 (2009).
61. Bullock, H. D., Hower, S. & Fields, K. A. Domain analyses reveal that *Chlamydia trachomatis* CT694 protein belongs to the membrane-localized family of type III effector proteins. *Journal of Biological Chemistry* **287**, 28078–28086 (2012).
62. Chen, Y. S. *et al.* The *Chlamydia trachomatis* Type III Secretion Chaperone Slc1 Engages Multiple Early Effectors, Including TepP, a Tyrosine-phosphorylated Protein Required for the Recruitment of CrkI-II to Nascent Inclusions and Innate Immune Signaling. *PLoS pathogens* **10**, e1003954 (2014).
63. Mueller, K. E., Plano, G. V. & Fields, K. A. New Frontiers in Type III Secretion Biology: the *Chlamydia* Perspective. *Infection and immunity* **82**, 2–9 (2014).
64. WHO simplified trachoma grading system. *Community eye health/International Centre for Eye Health* **17**, 68–68 (2004).
65. Vilella, A. J., Blanco-Garcia, A., Hutter, S. & Rozas, J. VariScan: Analysis of evolutionary patterns from large-scale DNA sequence polymorphism data. *Bioinformatics* **21**, 2791–2793 (2005).

## Acknowledgements

We are grateful to the study participants from The Gambia and Guinea-Bissau who volunteered to take part in the studies utilised in this manuscript. We thank Helen Fletcher and Martin Goodier (Infection and Immunology

Department, LSHTM) for helpful comments on the manuscript. This work was supported by the Wellcome Trust (grant number 097330/Z/11/Z [MJH, DCWM, RLB]) and a programme grant award from the Medical Research Council UK (DCWM, RLB). HP was supported by Bloomsbury Colleges (MJH, RDH). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Author Contributions

Conceptualisation: H.P., D.C.W.M., R.L.B., L.M.D.L.M., N.R.T., S.L.R.L., R.D.H., M.J.H. Fieldwork: N.F., H.J., P.M., E.C., M.N., A.L., S.E.B., M.J.H. Methodology: H.P., A.T., L.M.D.L.M., Ch.R., M.J.H. Investigation: H.P., A.T., N.F., N.R.T., M.J.H. Writing: H.P., M.J.H. Manuscript review: H.P., N.F., S.E.B., Ch.R., D.C.W.M., R.D.H., L.M.D.L.M., M.J.H.

### Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-09193-2](https://doi.org/10.1038/s41598-017-09193-2)

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017